

The Snowden Effect: Persistence and Geographical Reach

Ahmed Jellouli Anas Machraoui Karim Hadidane
ahmed.jellouli@epfl.ch anas.machraoui@epfl.ch karim.hadidane@epfl.ch

CS-401 Applied Data Analysis, EPF Lausanne, Switzerland

Abstract—Chilling Effects are manifestations of self-inhibiting behaviour out of fear of legal repercussions. Our project builds on previous works that studied the impact of the NSA/PRISM surveillance revelations on traffic to privacy-sensitive English Wikipedia articles. Here, we attempt to answer further research questions by first extending the temporal scope of the analysis to monitor the persistence of the chilling effect phenomenon. We then perform similar analysis on German Wikipedia articles in order to investigate the geographical span of this effect. We don't find any conclusive evidence of a persistent Chilling Effect on English Wikipedia articles. On the other hand, no statistically significant effect is found for the German pages.

I. INTRODUCTION

In [1], Jonathan W. Penney finds statistically significant evidence of a Chilling Effect on Wikipedia pageviews, arising after the 2013 PRISM/NSA revelations. The author uses English Wikipedia page views data, and interrupted time-series regression methods in order to demonstrate a clear impact of the Snowden revelations on the amount of views on privacy-sensitive articles, namely terrorism related pages. In this work, we try to answer two additional questions about this Chilling Effect. First, we study its persistence by gathering pageviews data extended until December 2015, and performing similar interrupted time series analysis (along with comparator groups), on the data. Second, we test the hypothesis of a chilling effect on German Wikipedia articles, using both a regression and matching approach.

II. PREVIOUS WORK

A. Chilling effects paper

In the paper “Chilling Effects: Online Surveillance and Wikipedia Use” [1], Jonathon W. Penney uses Wikipedia pageview counts to test the hypothesis of a Chilling Effect on English Wikipedia articles. For this, the author gathers data for privacy-sensitive articles (related the Department of Homeland Security keywords for Terrorism), and performs interrupted time series (ITS) analysis on the total aggregated monthly pageviews. ITS regression takes the form of:

$$y_t = \beta_0 + \beta_1 time + \beta_2 intervention + \beta_3 postslope \quad (1)$$

where y_t is the total monthly pageviews, $time$ is a discrete variable representing the month index from the start, $intervention$ in an indicator random variable (coded 1 after

the intervention), and $postslope$ is coded sequentially starting the intervention.

This analysis is widely used for studies related to policy impacts, and allows easy interpretation of the effects of an intervention.

The author finds a significant Chilling Effect on the terrorism related articles pageviews.

B. Facebook's MUSE embeddings

Another important work on which our project is built, are Facebook's *Multilingual Unsupervised and Supervised Embeddings* [2]. These are word embeddings for different languages, that are projected in the **same vector space**. This means, for example, that the embedding for the word "cat" is very close to that of "chat" in french.

These embeddings are useful when working with textual data from different languages, and allow us to discover cross-lingual semantic similarities.

III. DATA SCRAPING AND EXPLORATION

A. Data Scraping

Our original idea was to use the same dataset as used in [1], which was collected using `stats.grok.se`. However, the website has been down for a few weeks. Another option was to use Wikimedia's `pagecounts-raw` dataset. This alternative was, however, both time and space consuming and, thus, not feasible for the limited work time. Finally, we were able to find another source for the page views counts. This is a REST API provided at <http://petermeissner.de:8880> [3].

The API provides pageview data up until December 2015, and worked very well for English articles. However, there were problems for the German articles pageviews as the API did not return any data for many of them, even when the articles do exist on Wikipedia. As will be explained in section IV-B, we deal with this limitation by essentially downloading many German articles related to the same privacy-sensitive topics. We then find for each English article that does not have a perfect match, the 'closest' or 'best' match instead.

B. Data visualization

1) *Initial test*: We use the API in order to download the pageviews data for the 48 privacy-sensitive articles mentioned in [1], as well as their German counterparts, if this article

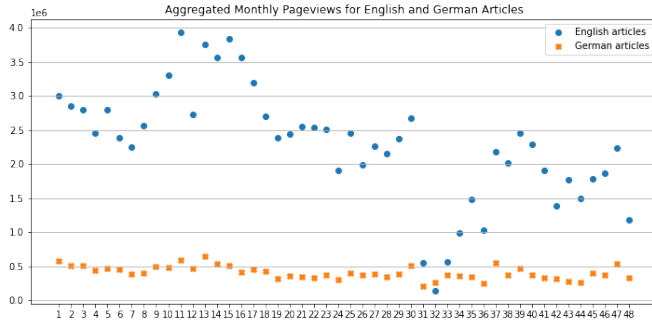


Fig. 1: Aggregated monthly pageviews for Jan 2012 until Dec 2015, for the English and German articles.

exists in German Wikipedia and its pageviews were available, which corresponded to a total of 40 German articles. Our initial idea to test the validity of the data obtained is to compare it with the dataset provided for the replication tasks. If pageview counts were to be very different, that would mean our source is not very reliable. Interestingly we found a *perfect match* in the day-to-day pageviews for both datasets, up until August 2014. This was a good indication that we had reliable data.

2) Scatter plots of the aggregated monthly pageviews:

For the same dataset, we plot the aggregated total monthly pageviews (figure 1). We notice that the amount of pageviews for the English articles is a far greater than that of its German counterpart, which is expected especially given that we have less articles. More importantly, some points stand out as outliers for the English Wikipedia pageviews, especially in months ranging from Jul 2014 until Dec 2014.

3) *Testing for outlier in English pageviews data:* Here, we look for articles that have zero pageviews reported for a certain month. We find **35** of these case for the English pages and **39** for the German ones. It is therefore important to keep this in mind when interpreting the results of our methods.

IV. METHODS AND RESULTS

A. Testing the Chilling Effects hypothesis for a longer time interval

With the page views data extended until December 2015, the permanent chilling effect hypothesis can be further tested. We fit the Segmented regression model on the data of page views of the terrorism-related articles from Jan 2012 to Dec 2015, with as an intervention at the time of the Snowden revelations. We use, for consistency, the 30 terrorism-related articles, chosen by the paper as the most privacy-sensitive ones after using crowdsourcing [1]. The model yields low p-values and *Adj R Squared* of 0.564.

We also studied the *two* Comparator groups used in the paper, the 23 domestic-security related articles and the 34 Infrastructure-security related articles. The regression model on the first group is statistically significant. The Adjusted R squared is high (0.554) and the p-values of the coefficients are low (less than 5% for the intervention and post slope

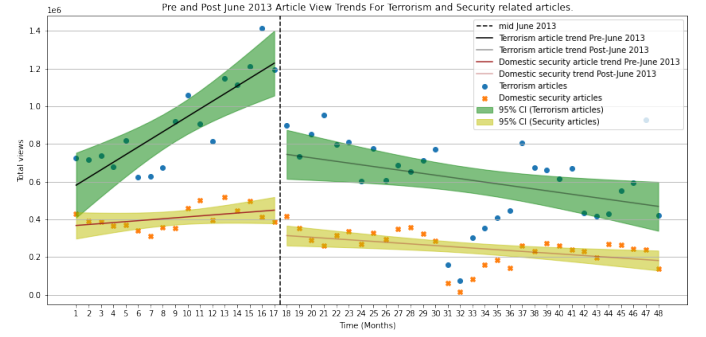


Fig. 2: ITS regression applied on the Terrorism-related vs Domestic-security articles

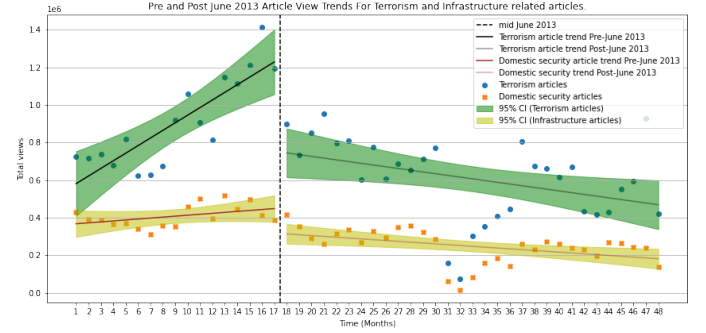


Fig. 3: ITS regression applied on the Terrorism-related vs Infrastructure-security articles

parameters). On the other hand, the results are not statistically significant when the model is applied on the second Comparator group with high p-values (higher than 5%).

Table 1 shows the relevant results of fitting the segmented regression model on the three groups of articles. The fitted regression lines for the three groups are also shown in Fig. 2. and Fig. 3. Therefore, we can not confirm the hypothesis that a permanent chilling effect on the Terrorism-related articles is taking place.

model fit on	Terrorism	Domestic	Infrastructure
Adj R sqr	0.524	0.554	0.552
p-values	0.000	0.005	0.068
	0.000	0.177	0.692
	0.000	0.022	0.482

TABLE I: Summary of fitting the segmented regression on Terrorism-related articles and the two Comparator groups

B. Testing the Chilling Effect hypothesis on German Wikipedia

1) *Initial regression:* As an initial step, we perform the same regression as done on the paper on the German article pageviews. For consistency with the paper [1], we discard the Hamas article for this task (Figure 4).

The model's fit was acceptable, we found an $Adj.R^2 = 0.444$ and $P > F = 0.0002$. However, the coefficients found for the `post_slope` and the `intervention`, although negative,

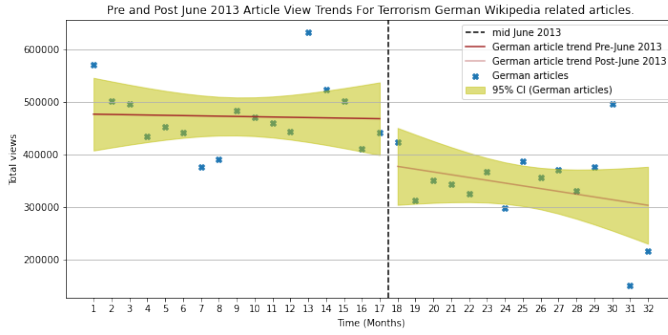


Fig. 4: ITS regression applied on the German articles, no statistically significant effect was found.

were not statistically significant. This is a first indication that there was no chilling effect on users of German Wikipedia.

2) *Matching of articles*: In order to confirm our initial result, we decided to perform matching on the German and English articles. Here we define:

- The **treatment**: The article belong to English Wikipedia.
- The **outcome**: There is a statistically significant lasting chilling effect on the pageviews of the article. This means that there is an immediate drop in pageviews after June 2013 ($p - value < 0.05$) and that there the slope of the regression is negative after the interruption ($p - value < 0.05$).

We match a slightly modified version of the 48 articles used in [1] to articles in German. Mainly, some of the pages used in [1] are 'disambiguation' pages. These cannot be easily matched to German versions. For this case, we simply match the page to the most privacy-sensitive link it leads to (chosen manually).

The matching is performed on the *semantic content* of the article. We use the following strategy for matching:

- If the article has a German equivalent Wikipedia page, it is matched to that page.
- Otherwise match the article to a 'close' German version.

For the second case, we use an Information-Retrieval approach:

- We extract many German articles corresponding to privacy sensitive Wikipedia categories. These categories were manually picked from categories of 'naturally matched' pages, i.e those who correspond directly to an English page. These articles are filetered so as to keep only those that have a considerable (≥ 10000) number of total pageviews over the considered period.
- For the unmatched pages and extracted German pages, we use an NLP pipeline on their summary. We perform (language-specific) tokenization, stop-words removal, lemmatization, and then map each article to the average of the Facebook MUSE embedding of its tokens. This pipeline is built using *spacy*.
- We use the *networkx* library in order to perform a maximum matching using the cosine similarity as a

proximity measure.

Having run this pipeline, we label resulting articles. This is done by performing regression on individual article pageviews, and conditioning on the significance and magnitude of the change after an intervention at June 2013.

Finally, we compute the Average Treatment Effect as:

$$ATE = \frac{1}{N} \sum_i^N y_{treat}^{(i)} - y_{-treat}^{(i)} \quad (2)$$

The result was an $ATE = 0.192$. This is significant evidence that English Wikipedia articles are more likely to be affected than their German counter-parts.

V. CONCLUSION

We find strong evidence that the Chilling Effect does not affect the German speaking public. However, although the ITS model for the extended period is significant, we cannot definitely conclude that the Chilling effect is permanent, as one of the comparator groups also shows a permanent decrease in pageviews.

REFERENCES

- [1] J. W. Penney, "Chilling effects: Online surveillance and wikipedia use," *Berkeley Technology Law Journal*, vol. 31, no. 1, pp. 117–182, 2016. [Online]. Available: <http://www.jstor.org/stable/43917620>
- [2] Facebook's muse embeddings. [Online]. Available: <https://github.com/facebookresearch/MUSE>
- [3] Wikipedia page view statistics late 2007 and beyond. [Online]. Available: <https://petermeissner.de/blog/2019/10/09/wikipediatrend-v2.1.4/>